Techniques for Improving the Labelling Process of Sentiment Analysis in the Saudi Stock Market

Hamed AL-Rubaiee
Department of Computer Science and Technology
University of Bedfordshire
Bedfordshire, United Kingdom

Khalid Alomar
Department of Information Systems
King Abdulaziz University
Jeddah, Kingdom of Saudi Arabia

Renxi Qiu
Department of Computer Science and Technology
University of Bedfordshire
Bedfordshire, United Kingdom

Dayou Li
Department of Computer Science and Technology
University of Bedfordshire
Bedfordshire, United Kingdom

***Abstract*—Sentiment analysis is utilised to assess users' feedback and comments. Recently, researchers have shown an increased interest in this topic due to the spread and expansion of social networks. Users' feedback and comments are written in unstructured formats, usually with informal language, which presents challenges for sentiment analysis. For the Arabic language, further challenges exist due to the complexity of the language and no sentiment lexicon is available. Therefore, labelling carried out by hand can lead to mislabelling and misclassification. Consequently, inaccurate classification creates the need to construct a relabelling process for Arabic documents to remove noise in labelling. The aim of this study is to improve the labelling process of the sentiment analysis. Two approaches were utilised. First, a neutral class was added to create a framework of reliable Twitter tweets with positive, negative, or neutral sentiments. The second approach was improving the labelling process by relabelling. In this study, the relabelling process applied to only seven random features (positive or negative): "earnings" (ارباح), "losses" (خسائر), "green colour" (باللون_الاخضر), "growing" (زياده), "distribution" (توزيع), "decrease" (انخفاض), "financial penalty" (غرامة), and "delay" (تاجيل). Of the 48 tweets documented and examined, 20 tweets were relabelled and the classification error was reduced by 1.34%.**

***Keywords*—Opinion mining; association rule; Arabic language; sentiment analysis; Twitter**

I. INTRODUCTION

Classifying a sentiment polarity as positive or negative is challenging due to the subjectivity factor of the sentiment polarity. Opinionated text can also carry some speech informality—such as sarcasm, subjective language, and emoticons—that makes the opinion detection harder. This required more understanding of the text beyond the facts being expressed [1]. In addition, sentiment polarity might contain positive and negative keywords that can make the labelling process unreliable. This occurred frequently in the neutral class, where one tweet might contain both positive and negative keywords. Labelling carried out by hand can cause human mislabelling sentiments. Therefore, adding the neutral class can help give flexibility for humans to have more options in the labelling process. However, this might cause less accuracy results. Since, the classifier techniques are used to cover the hall data set vectors, including the neutral data training. The reason behind this is that the data dictionary becomes larger which consist of all vectors that belong to positive, negative, and neutral. But humans will still label the data manually, which may create some mistakes in the labelling process. Consequently, the inaccurate classification creates the need to construct a relabelling process for Arabic tweets to remove noise on labelling. The main goal of the relabelling process is to remove the labelling noise. This will update experts' knowledge about labelling, which may lead to better classification. This is necessary because of the high degree of noise in labelling texts. Especially, for comments are long and consist of multiple sentences such as blogs [2].

This paper presents techniques for improving the labelling process of sentiment analysis. Section 2 shows the need to improve the labelling process for the neutral class. Section 3 demonstrates the Arabic sentiment analysis. Section 4 shows the experiment-classification into positive, negative, and neutral classes. Section 5 shows the need to improve the labelling process by relabelling. Section 6 demonstrates the process of relabelling. Section 7 analyses the experimental findings from Saudi stock market data. The final section contains a conclusion and recommendations for further work in this area.

II. IMPROVING THE LABELLING PROCESS WITH THE NEUTRAL CLASS

Researchers tend to ignore the neutral class under the hypothesis that there is less learning sentiment from neutral texts compared to positive or negative classes. The neutral class is useful, though, in real-life applications since sentiment is sometimes being neutral and excluding it forces instances into other classes (positive or negative) [3]-[5]. In addition, sentiment polarity might have positive and negative keywords that can make the labelling process unreliable. This happened regularly in the neutral class, where one tweet might have both positive and negative keywords. Labelling carried out by hand can cause human mislabelling of sentiments. Therefore, adding the neutral class can give humans more flexibility and options in the labelling process. However, this might cause less accurate results, since the data dictionary, which consists of all vectors that belong to positive, negative, and neutral, becomes larger.

III. Arabic Sentiment Analysis

Limited research has been conducted on Arabic sentiment analysis, so this is a field that is still in its early stages [6]. However, Boudad et al. [7], [8] reviewed the challenges and open issues that need to be addressed and explored in more depth to improve Arabic sentiment analysis, finding that these include domain, method of sentiment classification, data pre-processing, and level of sentiment analysis. They show that, in contrast to work on the English language, work on Arabic sentiment analysis is still in the early stages, and there are a lot of potential approaches and techniques that have not yet been explored. Another work carried out by Ibrahim et al. [9] have presented a multi-genre tagged corpus of MSA and colloquial language, with a focus on Egyptian dialects. Interestingly, they suggested that NLP supplements, which have been applied to other languages like English, are not valid for processing Arabic directly. Further, Abdulla et al. [10] explored the polarity of 2,000 collected tweets on various topics, such as politics and art. They used SVM, NB, KNN, and D-tree for their documents' sentiment classification. They showed that SVM and NB have better accuracy than other classifiers in a corpus-based approach. Their results reported that the average accuracy of SVM was 87.2%, while the average accuracy of NB was 81.3%. El-Halees's [11] combined approach classified documents using lexicon-based methods, used these as a training set, and then applied k-nearest neighbour to classify the rest of the documents.

IV. Experiment-Classification into Positive, Negative, and Neutral Classes

In this paper, Twitter has been chosen as a platform for opinion mining in trading strategy with the Saudi stock market to carry out and illustrate the relationship between Saudi tweets (standard and Arabian Gulf dialects) and the Saudi stock market index. The tweets' source data was obtained from the Mubasher company website in Saudi Arabia, which was extracted from the Saudi Stock Exchange (which is known by TASI[1] Index). This experiment will add the neutral class with the N-gram feature. For this study machine learning approach utilised, in which a set of data labelled as positive, negative. The classifiers, which were used to explore the polarity of all the classes' data was Naive-Bayes and SVM. Two different weighting schemes (Term Frequency–Inverse Document Frequency (TF-IDF) and Binary Term Occurrence (BTO)) were used for all classes (Positive, Negative, and Neutral). Table I shows the comparison between the classifiers with the neutral class in term of class accuracy, recall, and precision. Table I shows that SVM with TF-IDF worked better to classify the targeted documents when we add the neutral class.

TABLE I. Precision and Recall for Positive, Negative, and Neutral Classes using N-Gram Feature with SVM and NB

Class	Classifier Name	Weighting Schemes	Accuracy	Class Recall	Class Precision	Classification Error
All Classes	Naive-Bayes	BTO	74.16%	74.99%	74.01%	25.84%
		TF-IDF	72.05%	72.05%	71.51%	27.95%
	SVM	BTO	83.02%	81.14%	84.78%	16.98%
		TF-IDF	83.58%	81.67%	84.62%	16.42%

TABLE II. Precision and Recall For Positive, Negative, and Neutral Classes With SVM and Naive-Bays

Classifier	Accuracy	Recall	Precision
Naive-Bayes with BTO	74.16%	74.99%	74.01%
SVM with TF-IDF	83.58%	81.67%	84.62%

To sum up the classification experiment, the best accuracy achieved by SVM with TF-IDF was 83.58%. Moreover, the best recall and precision was achieved by SVM with less classification error. The analysis shows similar result for SVM with both schemas and only slight differences between recall and precision. Table II shows the comparison between the classifiers in terms of class accuracy, recall, and precision.

A one-to-one model shows the relationships between the positive, negative, and neutral classes and the TASI. The build model illustrates the results in sentiment analysis by showing the positive, negative, and neutral opinions as well as the TASI closing values. Fig. 1 shows the relation between labelling by human operators and the TASI for the Saudi stock market for positive negative and neutral classes between the middle of March 2015 to May 10, 2015. It can clearly be seen that the positive, negative, and neutral classes rise and fall with each other over time; the greatest score for neutral classes occurred on 21/4/2015; the lowest neutral class score occurred on 25/3/2015; and the lowest negative class score occurred on 28/4/2015. Only once did the neutral class get lower than the negative class, over a four-day period between 23/3/2015 and 26/3/2015. At that point, the TASI started to fall sharply. The neutral class frequently went above the positive class, but TASI remained the same. In conclusion, the neutral class mostly rose and fell with TASI. This indicates that the neutral class is an important consideration in the sentiment analysis process.

V. Improving the Labelling Process with Relabelling

High dimensionality in texts makes text pre-processing very significant in text classification problems, including sentiment analysis [12], [13]. This problem increases once the dimensionality becomes higher, like when adding neutral class for the classification. For example, in the previous experiment conducted to improve the labelling process of the neutral class, there was approximately 17% misclassification when SVM were used and approximately 28% misclassification when NB

[1] https://www.tadawul.com.sa

was used to classify the documents. In addition, labelling the documents conducted manually by humans may have introduced mistakes into the labelling process even when the neutral class was added. Thus, the inaccurate classification creates the need to construct a relabelling process for Arabic tweets to remove the noise on the original labelling. Below are the suggested steps for a relabelling process for Arabic sentiments analysis.

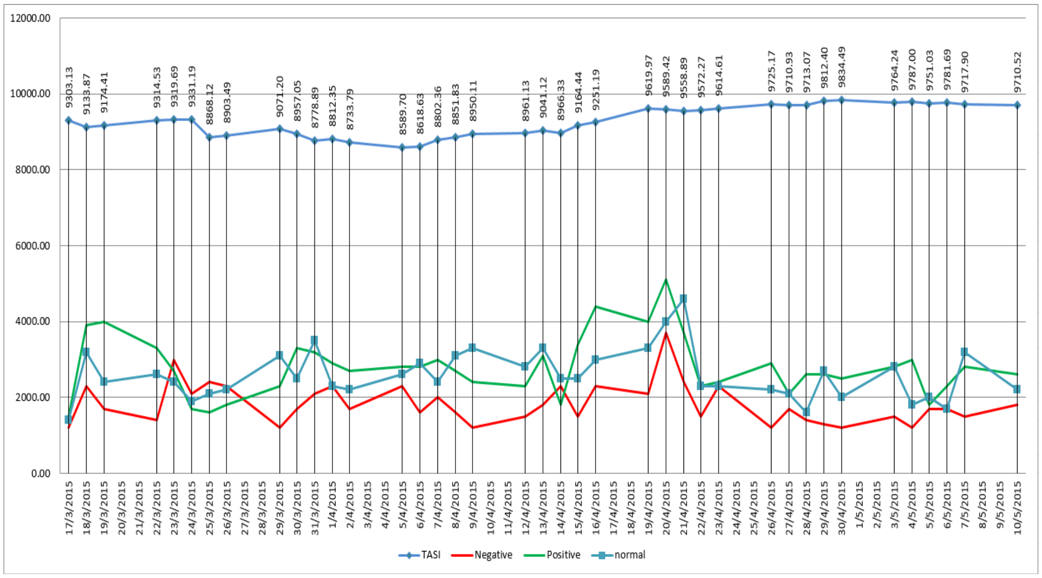

Fig. 1. The relation between labelling by human operators and the TASI for the Saudi stock market for positive, negative, and neutral classes.

The most challenging part of the process is feature selection since any feature can occur in all classes (positive, negative, and neutral). In addition, the main difficulty is to find out how many times the feature occurred in each class. Therefore, a wordlist technique was used to represent the text data as a matrix to show the frequent occurrence of each term within the three classes. Next, filtering the feature helps to select the highest features presented by the wordlist process. Then, in order to understand the sentences' structure and the sentiments behind them, a visualisation technique was used. This visualisation technique was applied to all data to achieve both a high level of understanding of the general structure and of the sentiment within an accumulated corpus. In other words, visualising the text shows the vital importance of the correlation between terms involved in the textual contents in general. However, visualisation shows only the feature with all the related terms in the textual contents without showing the classes they belong to. Using the wordlist technique with the visualisation can produce the important features created by the wordlist technique during the pre-process stage. Overcoming the visualisation limitation for the important features in the targeted text is essential to the relabelling process. After that, association rules extracted from the documents that have features that occurred in a questionable class. Association rules were generated regardless of the minimum support and minimum confidence threshold using the visualisation technique for the features that belong to the questionable class. By following these processes, the documents that have features in the questionable class can be relabelled again and the noise of the original labelling will be reduced.

VI. Process of Relabelling

Fig. 2 demonstrates the relabelling process for the Arabic sentiments analysis. The process started by collecting the corpus of data. The relevant training data were labelled and saved and the irrelevant training data were discarded.

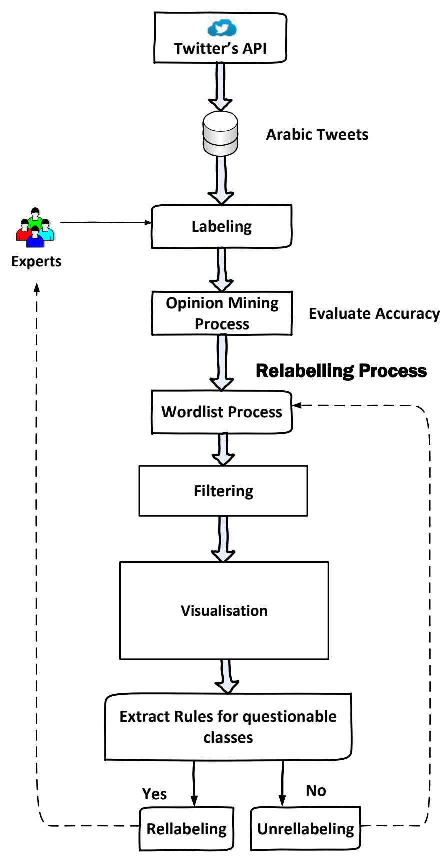

Fig. 2. Relabelling process.

The second step was to pre-processing the data by cleaning up hashtags, duplicate tweets, retweets, URLs, and special characters, and preparing the collected data for the labelling process. The next step is to label the cleaned data as positive, negative, or neutral by the expert in the domain. After that, the relabelling process consists of several steps: the Wordlist process, Filtering, Visualisation, Extract Rule, and Relabelling.

A. Wordlist Process

Fig. 3 shows the wordlist process. This phase uses the same corpus classified—positive, negative, or neutral—and the same data pre-processing procedure used in Opinion mining for the positive, negative, and neutral classes. The goal of visualising association rules as wordlists[2] is to have data sets that contain a row for each word and attributes for the word itself, and the number of labelled documents in each class for each term or word occurring in the training data. One other word represents the text data as a matrix to show the frequent occurrence of each term within the three classes. The key feature in this process was n-gram, which represents the correlation between the feature selection and other terms with their frequent occurrence for just two nods within the all-classes data.

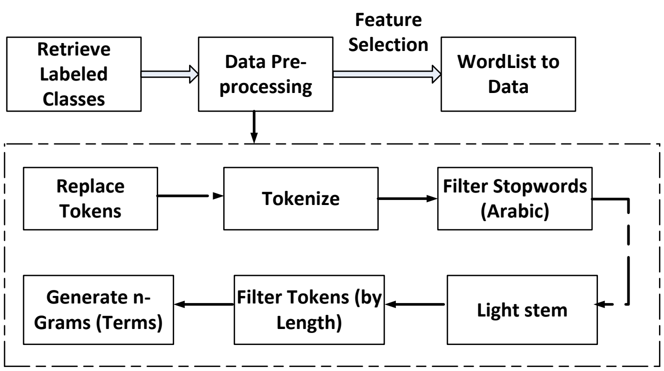

Fig. 3. Wordlist process.

B. Filtering

Features in the context of opinion mining are the words, phrases, or terms that strongly express the opinion as three polarities: positive or negative or neutral. In other words, features are the keywords chosen for the text as positive or negative. That means features have a higher impact on the orientation of a text than other words in the same texts. The impact of feature selection is to help to reduce the dimensionality of a text to increase the classification accuracy. Features in the text are considered explicit or implicit. Features appear in a text as explicit, whereas the feature does not appear is implied [14]. In the proposed process, the explicit features only considered.

C. Visualization

The importance of visualising text is to understand the sentences' structure and the sentiments behind them. Visualising the text shows the vital importance of the correlation between terms in the textual contents. The first step of the visualising technique is to produce the important features created by the wordlist process. Then, it was decided to select one of the features that appear in the dictionary created by the wordlist. Selection of the feature was done randomly to cover high-, average-, and low-frequency features to prove the concept of investigating the labelling noise. The next step is to visualise the selected feature in all-classes data as a wordlist representation. The wordlist shows how frequently the selected feature occurs in positive, negative, and neutral classes. If the selected feature was positive sentiment and occurred in other classes, such as neutral or negative, then the other classes (neutral, negative) are considered as a questionable class. In other words, if the selected feature is from the positive list, then it should occur only in the positive class—otherwise, this feature occurring in different classes would be considered a questionable class. A strong positive keyword should affect the text to be classified as positive unless there is a negation. Besides, it should not occur in the neutral class unless there are other words in the text affecting the sentiment. However, features that happened in a questionable class need further investigation to confirm the correctness of the labelling.

Fig. 4 illustrates the visualisation association rules process. In this phase, the same corpus classified as positive, negative, or neutral was used in this stage; and the same data pre-processing procedure used in the opinion mining process was carried out. After that, FP-Growth was used to discover frequent items discovery regarding the minimum support and minimum confidence threshold. Then, association rules were generated to expose the relationships between seemingly unrelated data. The output of visualization is the association of the high-frequency terms correlated with the selected feature presented previously from the wordlist process.

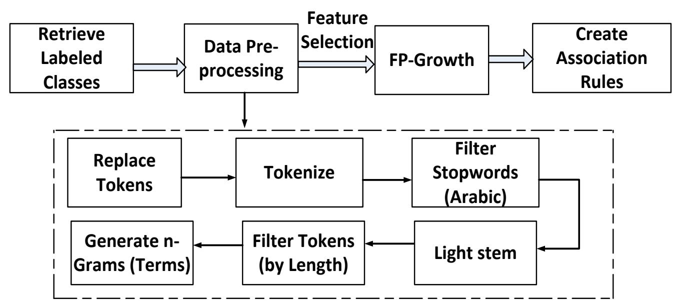

Fig. 4. The visualisation-generated association rules process.

D. Association Rules

The importance of association rule mining is to extract interesting correlations, frequent patterns, associations, or casual structures between sets of items in the transaction databases [15]. Association rule mining is divided into two steps. One, frequent patterns are mined about the threshold minimum support. Second, association rules are built according to the threshold minimum confidence [16]. Some terms or words appear with higher frequency in the dataset, while others rarely occur. In this case, the values of the minimum support will control the rule discovery. For instance, if the minimum support is set at a high value, rules that infrequently occur will not be found. Otherwise, if the minimum support is set at a low value, rules that frequently occur will be found. This cause rules with high confidence have very little support might be ignored [17], [18].

[2] https://rapidminer.com/

Text mining is defined as knowledge revelation from textual databases. Rules are created by analysing data for frequent if/then patterns. The frequent if/then patterns were mined utilizing methods such as the Apriori algorithm and the FP-Growth algorithm [19], [20]. However, for this study, the FP-Growth method was used to discover the frequent item set in the targeted document [21, 22]. Since, the main advantages of the FP-Growth are: passes only two times over data-set, no candidate generation, and compresses data-set [23].

E. Extract Rules

In this phase, the extraction of association rules from collection of documents was based on the features created by the wordlist. Association rules were generated around the minimum support and minimum confidence threshold using the previous visualisation process; the only difference here is the data we are going to use are the data belonging to the questionable class. This step focuses on extracting the rule that occurred less frequently in the questionable class within a specific document.

F. Relabelling

In this step, searching is the training data for the feature occurring less in each questionable classes according to the wordlist matrix. We ensured reliability of the relabelling applied by the expert for a specific document. Then, sentiment with labelling noise was sent as a recommendation to the expert to check its labelling.

VII. EXPERIMENT - RELABELLING

The Arabic text classifications regarding Saudi stock market opinions through the SVM algorithm were designed and implemented. The classification error was 16.42%. Therefore, a framework was created for relabeling.

The relabelling process started by representing the text data as a matrix to show the frequent occurrence of each term within the three classes. The relabelling process focused on representing the correlation between the feature selection and other terms with their high-frequency occurrence for just two nodes within the all-classes data. Table III shows the feature “earnings” (ارباح) as positive sentiment in the Saudi stock market domain. Table III shows the occurrence of the feature “earnings” (ارباح) in the positive, negative, and neutral classes.

TABLE III. OCCURRENCE OF THE FEATURE “EARNINGS” (ارباح)

Feature	Occurrence	Neutral	Positive	Negative
ارباح	304	14	223	67

Fig. 5 shows the association rules that related to the feature “earnings” (ارباح) in all classes with respect to the minimum support and minimum confidence threshold. The feature “earnings” (ارباح) entailed sharing the profits of some company in the Saudi stock market. Fig. 5 shows the most important rules for the feature “earnings” (ارباح) which is “earnings” --> “sharing out” [ارباح] --> [توزيع] (support: 0.031 confidence: 1), “earnings” --> “rise” [ارباح] --> [ارتفاع] (support: 0.071 confidence: 1), and “earnings” --> “decline” [ارباح] --> [تراجع] (support: 0.031 confidence: 1) The term (sharing out) [توزيع] correlated with the feature “earnings” [ارباح] to compose positive phrases distribute profits in the sentence. Further, the term “raise” [ارتفاع] correlated with the feature “earnings” [ارباح] to compose positive phrases “profits rises” in the sentence. On the other hand, the term “decline” [تراجع] correlated with the feature “earnings” [ارباح] to compose negative phrases profit “decline” in the sentence.

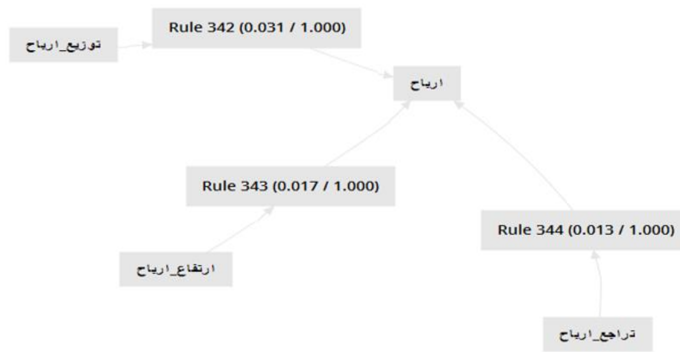

Fig. 5. Visualize the association rules for the feature “Earnings.(ارباح) ”

The next step is to find out from the wordlist representation the occurrence of the most frequent phrases that related to feature “earnings” (ارباح). Table IV shows that the phrase “high profits” [توزيع-ارباح] occurred 79 times—three in the neutral class, 66 in the positive class, and 10 in the negative class.

TABLE IV. PHRASES FOR THE FEATURE “EARNINGS” (ارباح) IN ALL DATA

Phrase - Terms	Occurrence	Neutral	Positive	Negative
توزيع-ارباح	79	3	66	10
ارتفاع-ارباح	37	0	37	0
تراجع-ارباح	31	0	0	31

From Table IV, the phrase “high profits” [ارتفاع-ارباح] does not need further investigation since it only occurred in the positive class. In addition, “profits decline” [تراجع-ارباح] does not need further investigation since it only occurred in the negative class. The phrase [توزيع-ارباح] occurred 79 times—three in the neutral class, 66 in the positive class, and 10 in the negative class. Since the phrase “distribute profits” [توزيع-ارباح] occurred in negative and neutral classes then both classes become questionable classes. Therefore, the feature “earnings” (ارباح) needs further investigation in order to find the association rules in both classes. As result, two scenarios will be followed: In the first scenario, the association rules that occurred for the feature “earnings” (ارباح) in the neutral class are extracted. Association rules are generated with regard to the minimum support and minimum confidence threshold using the previous process of the visualisation.

Fig. 6 shows that the feature “earnings” (ارباح) occurred with many rules that appeared in the premises column with the minimum support and minimum confidence values. However, according to the first scenario, the relevant rule here is [ارباح] --> [توزيع-ارباح] (support: 0.005 confidence: 1), which represents the phrase “distribute profits” [توزيع-ارباح] illustrated in the wordlist matrix in the neutral class.

Fig. 6. The correlation rules of the feature "Earnings" (ارباح) in neutral class.

From Fig. 6, the rule [ارباح] --> [توزيع-ارباح] (support: 0.005 confidence: 1) which represent the phrase "distribute profits" occurred in the neutral class. Therefore, the next step is to search for the phrase "distribute profits" [توزيع-ارباح] in the neutral class documents. Table V shows the phrase "distribute profits" [توزيع-ارباح] happened in three documents.

TABLE V. TERM "DISTRIBUTE PROFITS" [توزيع-ارباح] IN THE NEUTRAL CLASS DOCUMENTS

Original Labelling class	Original Arabic tweets with English translations	New Labelling class
Neutral	**تعلن** شركة اتحاد مصانع الأسلاك أسلاك عن تاريخ وطريقة **توزيع أرباح** النصف الثاني من عام 2014 م Union Wire Mills Company **announces** the date and method of **distributing dividends** for the second half of 2014	Positive
Neutral	**تعلن** شركة مكة للإنشاء والتعمير عن **توزيع أرباح** على المساهمين عن السنة المالية المنتهية في 1436/4/30هـ Makkah Construction & Development Company **announces** the **distributing dividends** to shareholders for the fiscal year ended 30/4/1436 H.	Positive
Neutral	**تعلن** شركة فواز عبدالعزيز الحكير وشركاه عن **توزيع أرباح** على المساهمين عن النصف الثاني من العام 2014 Fawaz Abdulaziz Al Hokair & Co. **announces** a **dividend distribution** for the second half of 2014	Positive

From Fig. 6, the rule [ارباح] --> [توزيع-ارباح] (support: 0.021 confidence: 1) which represents the phrase "distribute profits" occurred in the negative class. Therefore, the next step is to search for the phrase "distribute profits" [توزيع-ارباح] in the negative class documents. Table VI shows the phrase "distribute profits" [توزيع-ارباح] happened in 10 documents. Moreover, only one document has been found to satisfy the rule [ارباح] --> [توزيع-ارباح], so it has been sent again to the expert who labelled the document in the first stage. It can been seen from the structure in the rest of the nine documents that the phrase "distribute profits" توزيع أرباح has occurred with the negation [توزيع أرباح] --> [عدم], which is the right place for this term in the negative class.

TABLE VI. TERM "DISTRIBUTE PROFITS" [توزيع-ارباح] IN THE NEGATIVE CLASS DOCUMENTS

Original Labelling class	Original Arabic tweets with English translations	New Labelling class
Negative	**تأجيل** عمومية فيبكو المتضمنة الموافقة على **توزيع أرباح** Postpone the Vipco generality agreed on **dividend distribution**	Positive
Negative	مجلس إدارة العالمية للتأمين يوصي **بعدم توزيع أرباح** Global Insurance Board recommend on **not distribute dividend**	Negative
Negative	21 مايو عمومية الدرع العربي للموافقة على **عدم توزيع أرباح** May 21 general Arabian Shield for approval on **not distribute dividend**	Negative
Negative	غداً التصويت عن بُعد على بنود عمومية كيمانول والمتضمنة **عدم توزيع أرباح** Tomorrow remote voting on the general terms of Kimanol, including on **not distribute dividend**	Negative
Negative	ايس للتأمين تقر **عدم توزيع أرباح** وتنتخب أعضاء مجلس إدارتها ACE Insurance recognizes on **not distribute dividend** and elects its board of directors	Negative
Negative	تعلن شركة إتحاد إتصالات موبايلي عن توصية مجلس الإدارة **بعدم توزيع أرباح** عن الربع الأول من العام المالي 2015م Etihad Etisalat announces the recommendation of the Board of Directors on **not distribute dividend** for the first quarter of the fiscal year 2015	Negative
Negative	إدارة موبايلي تقر **عدم توزيع أرباح** عن الربع الأول من هذا العام Mobily's management confirms Directors on **not distribute dividend** for the first quarter of this year	Negative
Negative	9 يونيو عمومية موبايلي للموافقة على **عدم توزيع أرباح** June 9 Mobily's approval for **not distribute dividend**	Negative
Negative	609 % خسائر المتراكمة حتى 30 أبريل عمومية للموافقة على **عدم توزيع أرباح** 609% accumulated losses until 30 April general to approve on **not distribute dividend**	Negative
Negative	شركه صافولا توصي 22 يونيو عمومية للموافقة على **عدم توزيع أرباح** Savola recommends a June 22 general meeting to approve **not distribute dividend**	Negative

Fig. 7 shows that the feature "earnings" (ارباح) occurred with many rules that appeared in the premises column with the minimum support and minimum confidence values. According to the second scenario, the interested rule here is [ارباح] --> [توزيع-ارباح] (support: 0.021 confidence: 1), which represents the phrase "distribute profits" [توزيع-ارباح] illustrated in the wordlist matrix in the negative class.

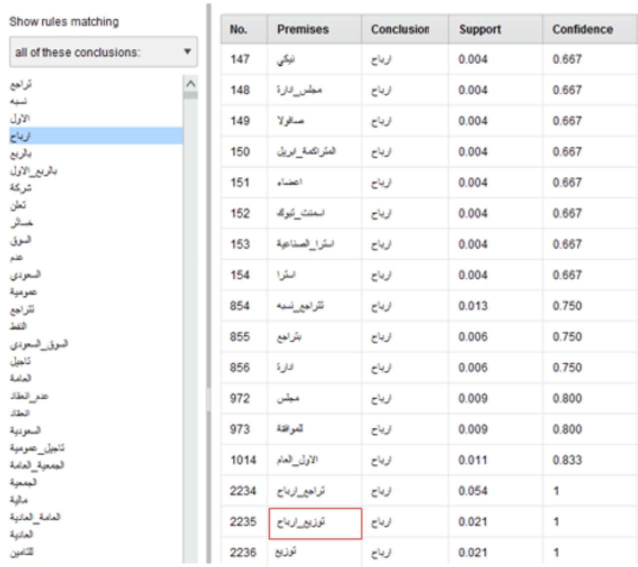

Fig. 7. The correlation rules of the feature "Earnings" in the negative class.

Fig. 8 shows the correlation rules that can happen with the feature "earnings" (ارباح) in the negative class. This process can identify the negation terms, such as عدم, which means the opposite of positive to help solving the negations problem with Arabic sentiment analysis.

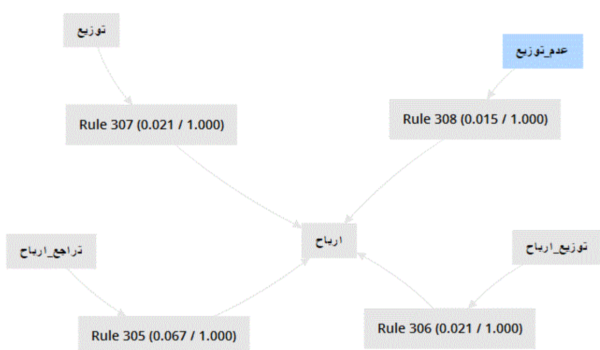

Fig. 8. The correlations of the "negation term" (عدم) and the feature "Earnings" (ارباح) in the negative class.

The second example verifies our experiment with another positive sentiment "Rise" (ارتفاع). Table VII shows the feature "Rise" (ارتفاع) as positive sentiment in the Saudi stock market domain. Table VII also shows the occurrence of the feature "Rise" (ارتفاع) in the positive, negative, and neutral classes.

TABLE VII. FEATURE "RISE" (ارتفاع) OCCURRENCE

Feature	Occurrence	Doc-tot	Neutral	Positive	Negative
ارتفاع	161	158	3	113	47

Fig. 8 shows the association rules related to the feature "Rise" (ارتفاع) in all classes with respect to the minimum support and minimum confidence threshold. The feature "Rise" (ارتفاع) is meant to obtain a financial advantage or benefit from an investment of some company in the Saudi stock market. In addition, Fig. 9 shows the most important rules for the feature "Rise" (ارتفاع), which is "earnings" --> "rise" [ارباح] --> [ارتفاع] (support: 0.091 confidence: 1), and the feature "percentage" --> "rise" [نسبه] --> [ارتفاع] (support: 0.013 confidence: 1). The term "earnings" [ارباح] correlated with the term "rise" (ارتفاع) to compose positive phrases High profits in the sentence. Further, the term "percentage" [نسبه] correlated with the feature "rise" (ارتفاع) to compose positive phrases high ratio in the sentence.

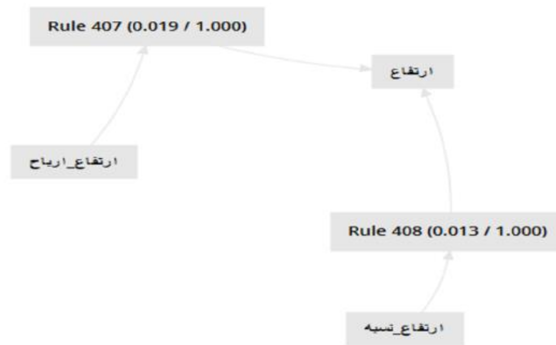

Fig. 9. Visualize the association rules for the feature "Rise" (ارتفاع).

The next step is to from the wordlist the representation of the occurrence of the most frequent phrases related to the feature "Rise" (ارتفاع). Table VIII shows that the phrase "high profits" [ارتفاع- نسبه-] occurred 30 times—zero in the neutral class, 23 in the positive class, and seven in the negative class.

TABLE VIII. PHRASES FOR THE FEATURE "RISE" (ارتفاع) IN ALL DATA

Phrase - Terms	Occurrence	Neutral	Positive	Negative
ارتفاع - ارباح	37	0	37	0
ارتفاع – نسبه	30	0	23	7

In Table VIII, the phrase "high profits" [ارتفاع-ارباح] does not need further investigation since it only occurred in the positive class. The phrase "high profits" [ارتفاع – نسبه] occurred 30 times—zero times in the neutral class, 23 in the positive class, and seven in the negative class. Since the phrase "high profits" [ارتفاع – نسبه] occurred in the negative class, the negative class becomes a questionable class. Therefore, the feature "Rise" (ارتفاع) in the negative class needs further investigation to find the association rules in the negative class. As a result, the second scenario will be followed: Extract the association rules that occurred for the feature "Rise" (ارتفاع) in the negative class. Association rules are generated around the minimum support and minimum confidence threshold using the previous process of the visualisation.

The feature "Rise" (ارتفاع) occurred with many rules that appeared in the premises column with the minimum support and minimum confidence values. However, according to the second scenario, the interested rule here is [ارتفاع] --> [ارتفاع – نسبه] (support: 0.015 confidence: 1) which represents the phrase "high profits" [ارتفاع – نسبه] illustrated in the wordlist matrix in the negative class. The rule [ارتفاع] --> [ارتفاع – نسبه] (support: 0.015 confidence: 1) which represents the phrase "high profits" [ارتفاع – نسبه] occurred in the negative class. Therefore, the next step is to search for the phrase "high profits" [ارتفاع – نسبه] in the negative class documents. Table IX shows the phrase "high profits" [ارتفاع – نسبه] happened in three documents. In addition, three documents have been found to satisfy the rule [ارتفاع] --> [ارتفاع – نسبه], so it has been sent again to the expert who labelled the document in the first stage. It can be seen from the structure of the three documents that the Arabic language is a derivative language in which new words are created from other words. For example, ارتفاع can create ارتفع, ترتفع, يرتفع, مرتفعا — all those words have the same meaning which "rise" in English language, and each term has different shape and this one of the problems that can be solved by the light stemming in the pre-process phase. Moreover, the replace token operator was used to replace "%" by the term "percentage" (نسبه) during the Pre-process phase. For this reason, the term "percentage" (نسبه) cannot be seen in the original documents.

TABLE IX. Phrase "High Profits" [ارتفاع-نسبة] in the Negative Class Documents

Original Labelling Class	Original Arabic Tweets with English Translation	New Labelling Class
Negative	صافي أرباح قطاع التجزئة **ترتفع** 12**%** رغم **تراجع** ثلث شركاته Net profit for the retail sector **increase** 12**%** despite a third of its companies **falling**	Positive
Negative	السوق السعودي **يرتفع** 04**%** بنهاية الأسبوع وصافولا تتراجع بقطاعها The Saudi market **is up** 04% at the end of the week and Savola is **falling** in its sector	Positive
Negative	المؤشر العام يغلق **مرتفعاً** وميدغلف **يتراجع** بأكثر من 8% General index closes **higher** and Medgulf **went down** more than 8%	Positive

For the third example for "losses" (خسائر), which is negative sentiment, first we visualize the most important rules for the term "losses" (خسائر). Table X shows the feature "losses" (خسائر) as negative sentiment in the Saudi stock market domain. Table X shows the occurrence of the feature "losses" (خسائر) in the positive, negative, and neutral classes.

TABLE X. Phrases for the Feature "Losses" (خسائر) in all Data

Feature	Occurrence	Neutral	Positive	Negative
خسائر	87	0	25	62

Fig. 10 shows the association rules related to the feature "losses" (خسائر) in all classes with respect to the minimum support and minimum confidence threshold. The feature "losses" (خسائر) means losing an investment in some company in the Saudi stock market. In addition, Fig. 10 shows the most important rules for the feature "losses" (خسائر), which is [خسائر] --> [المتراكمة] (support: 0.013 confidence: 1). The term "losses" [خسائر] correlated with the term "accumulated" (المتراكمة) to compose the positive phrase "accumulated losses" in the sentence.

Fig. 10. Visualize the association rules for the feature "losses" (خسائر).

The next step is to find out from the wordlist representation of the occurrence of the most frequent phrases that related to feature "losses" (خسائر). Table XI shows that the phrase "accumulated losses" [خسائر-المتراكمة] occurred 14 times—zero in the neutral class, four in the positive class, and 10 in the negative class.

TABLE XI. Phrases for the Feature "Accumulated Losses" (خسائر-المتراكمة) in all Data

Phrase -Terms	Occurrence	Neutral	Positive	Negative
خسائر-المتراكمة	14	0	4	10

Since the phrase "accumulated losses" [خسائر-المتراكمة] occurred in the positive class, the positive class becomes the questionable class. Therefore, the feature "losses" (خسائر) in the positive class needs further investigation in order to find out the association rules in the positive class. As result, the second scenario will be followed to extract association rules that occurred for the feature "losses" (خسائر) in the positive class. Association rules are generated with regard to the minimum support and minimum confidence threshold using the previous process of the visualisation. The rule [خسائر] --> [خسائر-المتراكمة] (support: 0.005 confidence: 1) which represent the phrase "accumulated losses" [خسائر-المتراكمة] occurred in the positive class. Therefore, the next step is to search for the phrase "accumulated losses" [خسائر-المتراكمة] in the positive class documents. Table XII shows the phrase "accumulated losses" [خسائر-المتراكمة] happened in three documents. Three documents have been found to satisfy the rule [خسائر] --> [[خسائر-المتراكمة, so it has been sent again to the expert who labelled the document in the first stage. It can be seen from the structure of the three documents that the term "descent" (إنخفاض) as negative sentiment came before the phrase "accumulated losses" [خسائر-المتراكمة], also a negative phrase that puts the three documents in the positive situation.

TABLE XII. Phrase "Accumulated Losses" [خسائر-المتراكمة] in the Positive Class Documents

Original Labelling Class	Original Arabic Tweets with English Translations	New Labelling Class
Positive	**تعلن** الشركة السعودية الهندية للتامين التعاونيه وفا للتامين عن **إنخفاض خسائرها المتراكمة** إلى أقل من 50 % من رأسمالها Saudi Indian Cooperative Insurance Company (WAFA) **announces** its **decrease accumulated losses** to less than 50% of its capital	Positive
Positive	**انخفاض الخسائر المتراكمة** لـ وفا للتأمين عن 50% من رأسمالها Wafa Insurance's **decrease accumulated loss** of 50% of its capital	Positive
Positive	تعلن الشركة المتحدة للتأمين التعاوني عن **إنخفاض خسائرها المتراكمة** إلى أقل من 50 % من رأسمالها United Cooperative Insurance Company announces a **decrease** in its **accumulated losses** to less than 50% of its capital	Positive
Positive	**انخفاض الخسائر المتراكمة** لـ وفا للتأمين عن 50% من رأسمالها **Decrease** loss of Wafa **Insurance accumulated** 50% of its capital	Positive

Finally, colour is used as for both positive and negative sentiment in this domain. For instance, green colour indicates a positive sentiment in the stock market domain, while red indicates negative sentiment with the HMI field in computing. Table XIII shows the feature “green” (الاخضر) as positive sentiment in the Saudi stock market domain. Table XIII also shows the occurrence of the feature “green” (الاخضر) in the positive, negative, and neutral classes.

TABLE XIII. PHRASES FOR THE FEATURE “GREEN” (الاخضر) IN ALL DATA

Feature	Occurrence	Neutral	Positive	Negative
الاخضر	15	0	10	5

Fig. 11 shows the association rules related to the feature “green” (الاخضر) in all classes with respect to the minimum support and minimum confidence threshold. The feature “green” (الاخضر) indicates that the Saudi stock market values are closing green. Fig. 11 shows the most important rules for the feature “green” (الاخضر), which is [الاخضر] --> [باللون] (support: 0.006 confidence: 1). The feature “green” (الاخضر) correlated with the term “colour” (اللون) to come up with the positive phrase “green colour” in the sentence.

Fig. 11. Visualize the association rules for the feature “Green” (الاخضر).

The next step is to find out from the wordlist representation of the occurrence of the most frequent phrases that related to the feature “green” (الاخضر). Table XIV shows that the phrase “green colour” [باللون_الاخضر] occurred 12 times—zero in the neutral class, eight in the positive class, and seven in the negative class.

TABLE XIV. PHRASES FOR THE FEATURE “GREEN COLOR” (باللون_الاخضر) IN ALL DATA

Phrase -Terms	Occurrence	Neutral	Positive	Negative
باللون_الاخضر	14	0	8	5

Since the phrase “green colour” [باللون_الاخضر] occurred in the negative class, the negative class becomes the questionable class. Therefore, the feature “green” (الاخضر) in the negative class needs further investigation in order to find out the association rules in the positive class. As result, the second scenario will be followed: Extract the association rules that occurred for the feature “green” (الاخضر) in the negative class. Association rules are generated with regard to the minimum support and minimum confidence threshold using the previous process of the visualisation. The rule [الاخضر] --> [باللون_الاخضر] (support: 0.011 confidence: 1) which represents the phrase “green colour” [باللون_الاخضر] occurred in the negative class. Therefore, the next step is to search for the phrase “green colour” [باللون_الاخضر] in the negative class documents. Table XV shows the phrase “green colour” [باللون_الاخضر] happened in seven documents. Seven documents have been found to satisfy the rule [الاخضر] --> [باللون_الاخضر], so it has been sent again to the expert who labelled the document in the first stage. It can be seen from the structure of the seven documents that the term “decline” (تراجع) as negative sentiment came sometimes before and after the phrase “green colour” [باللون_الاخضر], which is a negative term that puts the seven documents in the unreliable situation during the labelling process.

TABLE XV. PHRASE “GREEN COLOR” [باللون_الاخضر] IN THE NEGATIVE CLASS DOCUMENTS

Original Labelling Class	Original Arabic Tweets with English Translations	New Labelling Class
Negative	المؤشر العام يغلق باللون الأخضر والتطوير العقاري يتراجع 3% Index closes in green and real estate development falls 3%	Positive
Negative	الأسواق الخليجية تتراجع على خلفية عاصفة الحزم وأبوظبي باللون الأخضر Gulf markets retreat against the backdrop of the Al-Hazm Storm and Abu Dhabi in green	Negative
Negative	السوق السعودي يغلق متراجعاً بنسبة 17% وقطاع واحد باللون الأخضر The Saudi market closed down 17% and one sector in green	Negative
Negative	عمليات جني ارباح تغلق السوق السعودي دون 9750 نقطة وقطاعاً واحداً باللون الأخضر Profits taking closes the Saudi market below 9750 points and one sector in green	Negative
Negative	السوق السعودي يغلق باللون الأحمر وقطاع واحد باللون الأخضر The Saudi market closes in red and one sector in green	Negative

In addition, the same process was carried out for random features (positive or negative), namely, “growing” (زياده), “distribution” (توزيع), “decrease” انخفاض, “financial penalty” غرامة, and “delay”تاجيل . The result shows that, after completing the process, 23 documents were sent to experts to check the labelling. Of the 48 tweet documents examined, 20 tweets were relabelled.

In the last stage, the original data were updated according to the new labelling. Then, the updated data were loaded to run a new classification process. A comparison was carried out between the original classification and the new classification. Tables XVI and XVII show the performance accuracy for the SVM with TF-IDF schema for both, the original classification and the new classification, respectively. For the neutral class, the precision for the original classification is 92.59%, which rose to 94.09% for the new classification after the relabeling process. On the other hand, the recall for the original classification for the neutral class is 82.74%, which rose to 84.40% for the new classification after the relabeling process.

TABLE XVI. ALL CLASS PERFORMANCE ACCURACY IN ORIGINAL CLASSIFICATION FOR SVM WITH TF-IDF SCHEMA

	True Normal	True Positive	True Negative	Class Precision
Pred. Neutral	537	26	17	92.59%
Pred. Positive	103	757	119	77.32%
Pred. Negative	9	45	330	85.94%
Class Recall	82.74%	91.43%	70.82%	

TABLE XVII. ALL CLASS PERFORMANCE ACCURACY IN NEW CLASSIFICATION FOR SVM WITH TF-IDF SCHEMA

	True Normal	**True Positive**	**True Negative**	**Class Precision**
Pred. Neutral	541	21	13	94.09%
Pred. Positive	91	786	123	78.60%
Pred. Negative	9	36	323	87.77%
Class Recall	84.40%	93.24%	70.37%	

Table XVIII shows the comparison carried out between the result with the original classification and the new classification. The result showed that there was an improvement of 1.34% using SVM with TF-IDF with the new classification.

TABLE XVIII. ALL CLASS PERFORMANCE ACCURACY COMPARISON FOR SVM WITH TF-IDF SCHEMA

Classifier	**Accuracy**	**Recall**	**Precision**	**Classification Error**
SVM with the Original Classification	83.58%	81.67%	84.62%	16.42%
SVM with New Classification	84.92 %	82.66%	85.40%	15.08%

To sum up, the results show that our process can readily classify Arabic tweets. Furthermore, they can handle many antecedent text association rules for the positive class, the negative class, and the neutral class. The analysis shows the importance of the neutral class in sentiment analysis of Arabic documents; adding the neutral class shows different results of classification accuracy. The reason results are different is that the new vectors dictionary for the text data consists of all the words that belong to positive and negative classes as well. The obtained results help to understand the text structure and the sentiment behind them. Finally, these efforts are meant to add to the breadth of expert knowledge in this field and to be beneficial to the future of machine learning methods.

VIII. CONCLUSION

This study presents a relabeling process to enhance the classification accuracy and update the expert knowledge in the original labelling. Since human error occurs in labelling data, visualisation of the text can show the importance of the correlation between terms involved in the textual structured contents. This is especially apparent in the wordlist and the N-gram steps of the pre-process stage. After the relabeling process applied for random only seven features (positive or negative), namely, "earnings" (ارباح), "losses" (خسائر), "Green color" [باللون_الاخضر] , "growing" (زياده), "Distribution" (توزيع), "Decrease" انخفاض, "Financial penalty"غرامة , and "delay"تاجيل . Of the 48 tweets documented and examined, 20 tweets were relabelled and the classification error was reduced by 1.34%. The current study should be repeated in other domains such as education.

REFERENCES

[1] Haddi, E., Sentiment analysis: text, pre-processing, reader views and cross domains. 2015, Brunel University London.

[2] Yoshida, S., et al. Sentiment analysis for various SNS media using Naïve Bayes classifier and its application to flaming detection. in Computational Intelligence in Big Data (CIBD), 2014 IEEE Symposium on. 2014. IEEE.

[3] Go, A., R. Bhayani, and L. Huang, Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 2009. 1(2009): p. 12.

[4] Koppel, M. and J. Schler, The importance of neutral examples for learning sentiment. Computational Intelligence, 2006. 22(2): p. 100-109.

[5] Agarwal, A., et al. Sentiment analysis of twitter data. in Proceedings of the Workshop on Languages in Social Media. 2011. Association for Computational Linguistics.

[6] Hamed, A.-R., R. Qiu, and D. Li. Analysis of the relationship between Saudi twitter posts and the Saudi stock market. in Intelligent Computing and Information Systems (ICICIS), 2015 IEEE Seventh International Conference on. 2015. IEEE.

[7] Boudad, N., et al., Sentiment analysis in Arabic: A review of the literature. Ain Shams Engineering Journal, 2017: p. 1-12.

[8] Kaseb, G.S. and M.F. Ahmed, Arabic Sentiment Analysis approaches: An analytical survey.

[9] Ibrahim, H.S., S.M. Abdou, and M. Gheith. MIKA: A tagged corpus for modern standard Arabic and colloquial sentiment analysis. in Recent Trends in Information Systems (ReTIS), 2015 IEEE 2nd International Conference on. 2015. IEEE.

[10] Abdulla, N.A., et al. Arabic sentiment analysis: Lexicon-based and corpus-based. in Applied Electrical Engineering and Computing Technologies (AEECT), 2013 IEEE Jordan Conference on. 2013. IEEE.

[11] El-Halees, A.M., Arabic text classification using maximum entropy. IUG Journal of Natural Studies, 2015. 15(1).

[12] Das, S.R. and M.Y. Chen, Yahoo! for Amazon: Sentiment extraction from small talk on the web. Management science, 2007. 53(9): p. 1375-1388.

[13] Sebastiani, F., Machine learning in automated text categorization. ACM computing surveys (CSUR), 2002. 34(1): p. 1-47.

[14] Liu, B., Sentiment Analysis and Subjectivity. Handbook of natural language processing, 2010. 2: p. 627-666.

[15] Agrawal, R., T. Imieliński, and A. Swami. Mining association rules between sets of items in large databases. in Acm sigmod record. 1993. ACM.

[16] Wu, X., C. Zhang, and S. Zhang, Efficient mining of both positive and negative association rules. ACM Transactions on Information Systems (TOIS), 2004. 22(3): p. 381-405.

[17] Liu, B., Y. Ma, and C.K. Wong. Improving an association rule based classifier. in European Conference on Principles of Data Mining and Knowledge Discovery. 2000. Springer.

[18] Karthikeyan, T. and N. Ravikumar, A survey on association rule mining. International Journal of Advanced Research in Computer and Communication Engineering, 2014. 3(1): p. 2278-1021.

[19] Deng, Z.-H., K.-H. Luo, and H.-L. Yu, A study of supervised term weighting scheme for sentiment analysis. Expert Systems with Applications, 2014. 41(7): p. 3506-3513.

[20] Nasukawa, T. and J. Yi. Sentiment analysis: Capturing favorability using natural language processing. in Proceedings of the 2nd international conference on Knowledge capture. 2003. ACM.

[21] Han, J., J. Pei, and Y. Yin. Mining frequent patterns without candidate generation. in ACM sigmod record. 2000. ACM.

[22] Witten, I.H., et al., Data Mining: Practical machine learning tools and techniques. 2016: Morgan Kaufmann.

[23] Verhein, F., Frequent Pattern Growth (FP-Growth) Algorithm. School of Information Studies, The University of Sydney, Australia, 2008: p. 1-16.